Argonne NATIONAL LABORATORY

Accelerating and Benchmarking GAMESS quantum chemistry code on Mira

Maricris Mayes, Graham D. Fletcher, Yuri Alexeev
Argonne National Laboratory
Leadership Computing Facility

Mark S. Gordon
Iowa State University

U.S. DEPARTMENT OF ENERGY

Outline

- Introduction
  - Motivation
  - Overview of ab initio MD and quantum chemistry
  - GAMESS
  - Fragment Molecular Orbital
- Performance on Mira
- What is enabled by Mira over Intrepid?
- Code Modifications, New Algorithm: RATTLE

Towards High-Accurate Predictions of Bulk Properties of Water

Motivation:

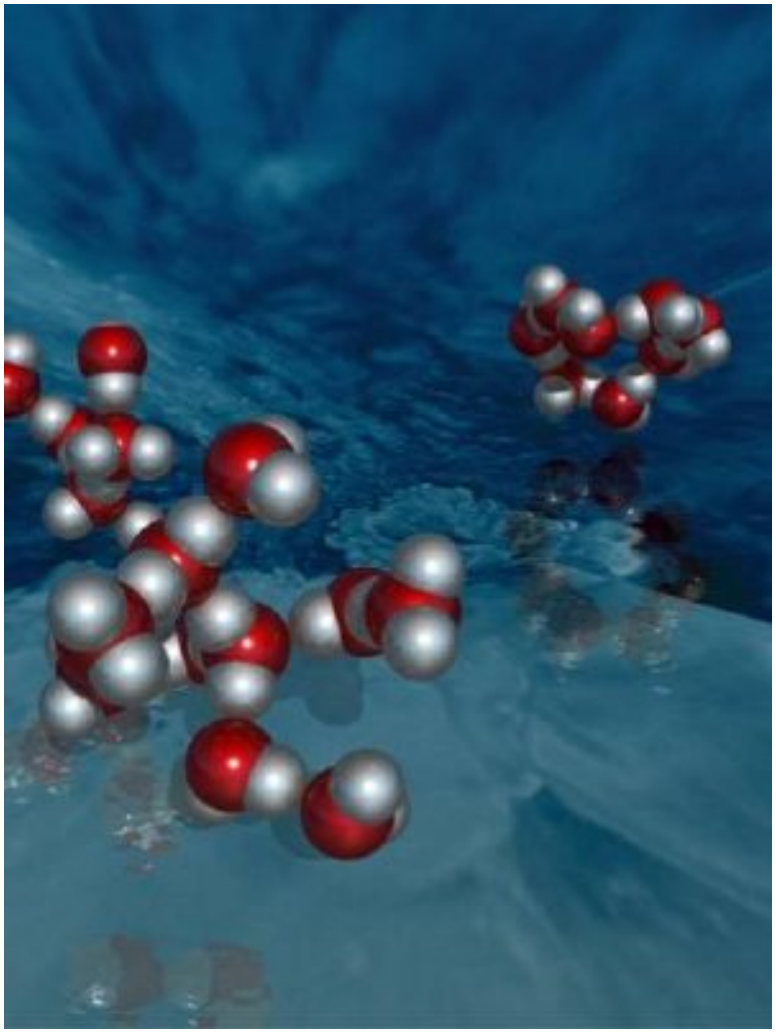
- the current state of theory and modeling of water is insufficient
- structure and dynamics of the H-bonding network in water has not been resolved

Model	Dipole Moment	Dielectric Constant	Self-Diffusion, 10^{-5} cm^2/s	Average Configurational Energy, kJ/mol	Density Maximum, °C	Expansion Coefficient, 10-4 $°C^{-1}$
TIP3P	2.35	82	5.19	-41.1	-91	9.2
TIP4P	2.18	52	3.29	-41.8	-25	4.4
Expt	2.95	78.4	2.30	-41.5	+3.984	2.53

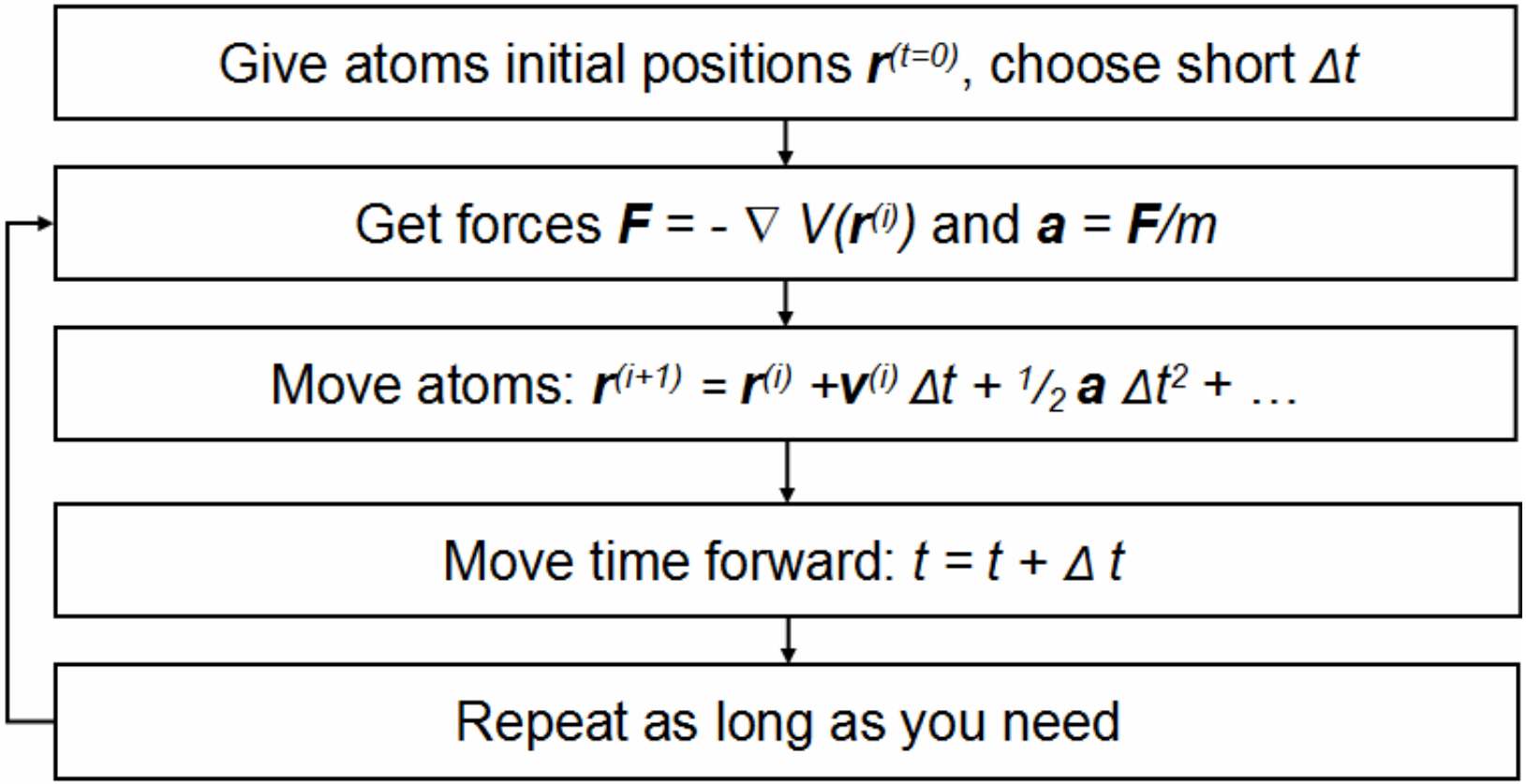
Towards High-Accurate Predictions of Bulk Properties of Water

Plan:

- perform *ab initio* molecular dynamics simulations of water using FMO2-MP2 to provide a comprehensive and quantitative comparison with experimental results and to calibrate empirical potentials



Molecular Dynamics



Ab Initio Quantum Chemistry

- Schrödinger Equation:

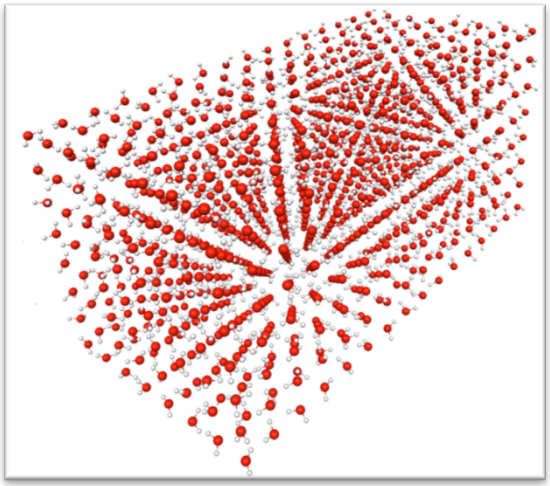
$$H_e \, \Psi_\mu(\mathrm{X};\mathrm{R}) = E_\mu(R) \, \Psi_\mu(\mathrm{X};\mathrm{R})$$

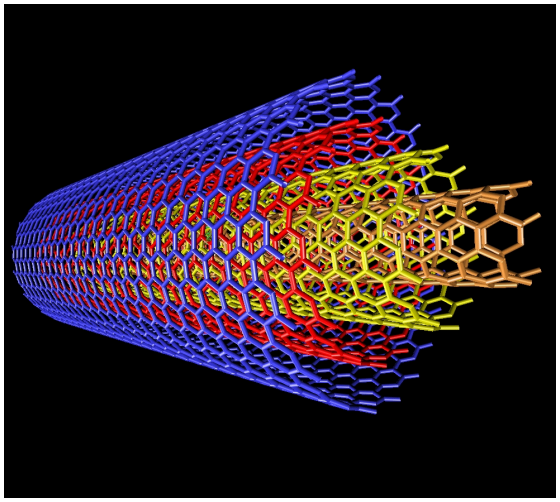
$$H_e = Z + V = \sum_{i=1}^{N} z(\mathrm{x}_i) + \sum_{i>j=1}^{N} v(\mathrm{x}_i, \mathrm{x}_j)$$

$$z(\mathrm{x}_i) = \frac{1}{2}\Delta_i + \sum_{A=1}^{M} \frac{Z_A}{R_{A_i}}, \qquad v(\mathrm{x}_i, \mathrm{x}_j) = \frac{1}{r_{ij}}$$

Major Challenge of Modern Quantum Chemistry

- To apply to systems with thousands of correlated electrons and basis functions





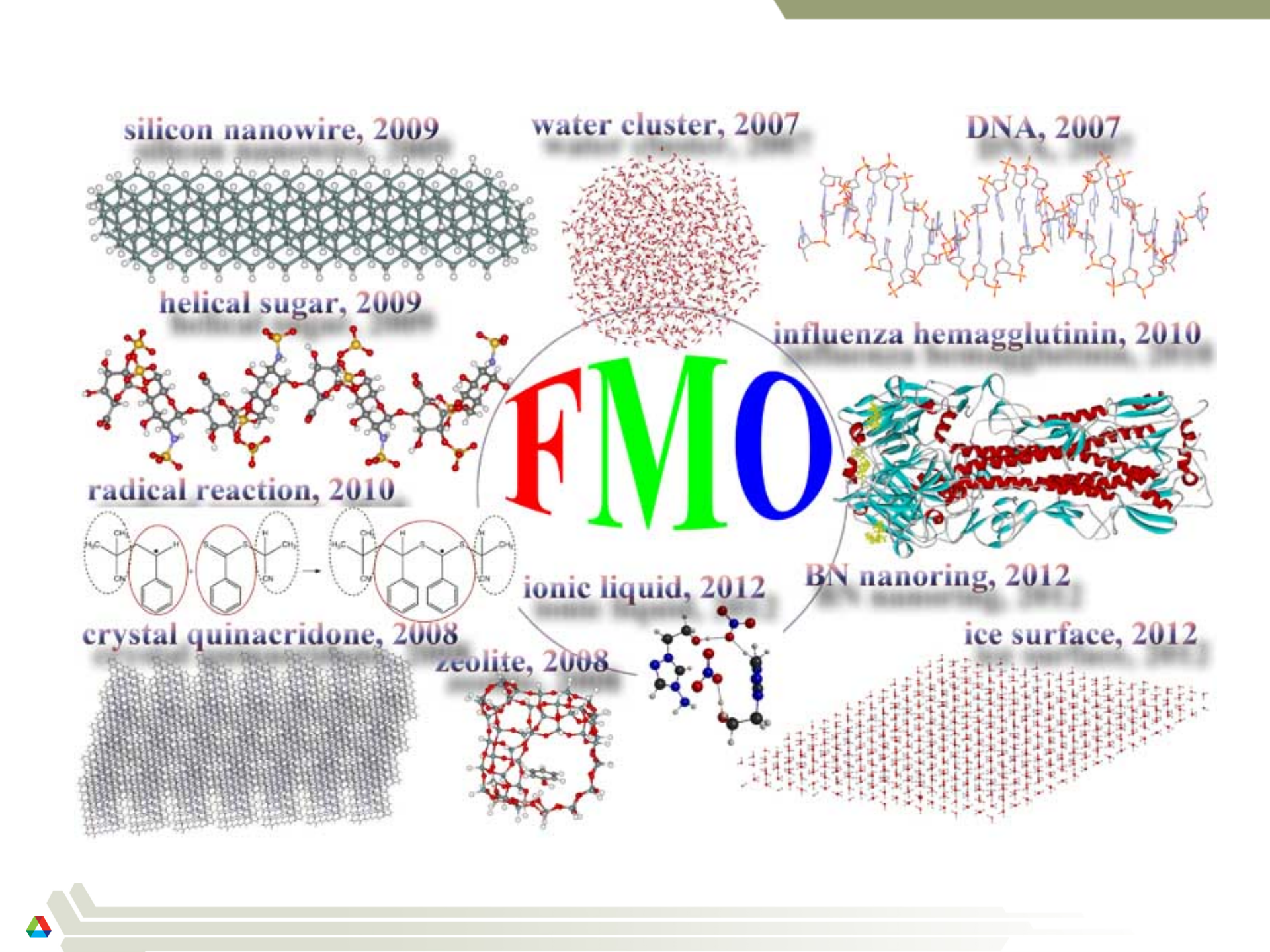
Major Challenge of Modern Quantum Chemistry

- In order to make QM computationally tractable to large systems:
  1. Availability of supercomputers
  2. Development of novel and scalable methods (e.g. Fragment Molecular Orbital method)

GAMESS

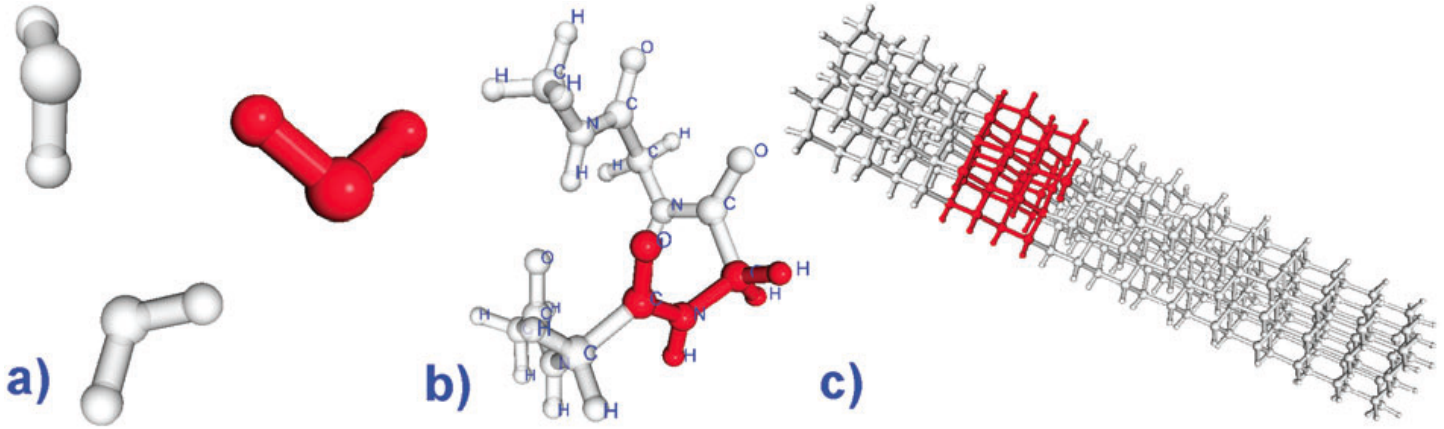
(General Atomic and Molecular Electronic Structure System)

- *Ab initio* quantum chemistry package
- Maintained by the research group of Prof. Mark Gordon at Iowa State University
- Implemented all major quantum mechanical methods
  - Hartree- Fock
  - Møller Plesset perturbation theory
  - Coupled-cluster
  - Multiconfiguration consistent field
  - Configuration interaction
  - Density functional theory
- Ported to all major architectures
- Free and widely used on everything from laptops to supercomputers
- About a million lines of code, with an associated parallelization library comprising 15,000 lines
- Highly scalable, including many distributed data algorithms


silicon nanowire, 2009
water cluster, 2007
DNA, 2007
helical sugar, 2009
influenza hemagglutinin, 2010
FMO
radical reaction, 2010
BN nanoring, 2012
ionic liquid, 2012
crystal quinacridone, 2008
zeolite, 2008
ice surface, 2012

Fragment Molecular Orbital, basics

- FMO in GAMESS is due to Fedorov, Kitaura
- Split system into **fragments** and solve for the electronic structure of each fragment in the **coulomb field** of the whole system
- How? Quantum effects are short-range
  - Long-range interactions can be described by the coulomb operator (point charges)



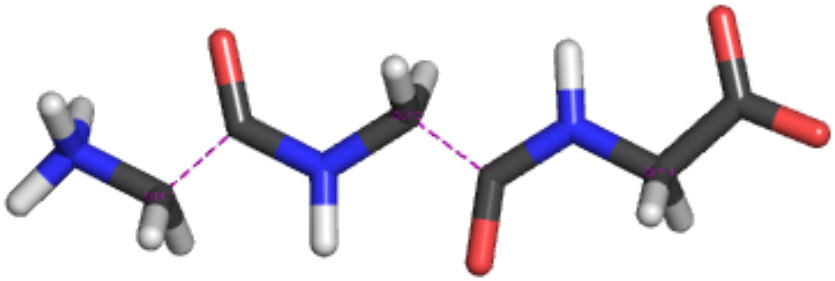
Fragment Molecular Orbital, basics

- Address major **bottlenecks** of quantum chemistry: Computational cost, memory, disc, bandwidth
- Main application to **proteins** (fragment: amino acid)
- Fully *ab initio* (no fitted parameters)
- Any QC method (MP2, MCSCF, CC, HF, DFT, …)

- Linear cost
- Highly scalable (coarse, fine-grained parallelism)

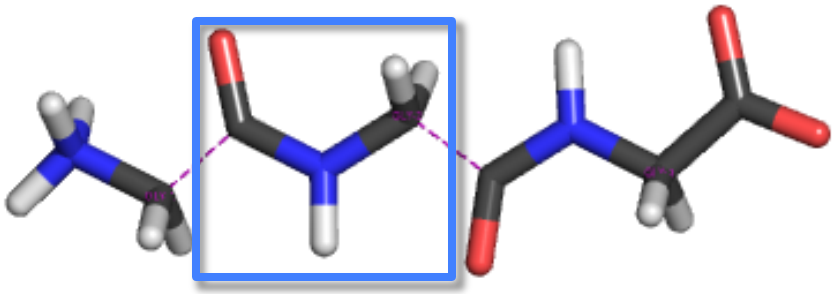
- Pair Interaction Energy Decomposition Analysis (PIEDA): Electrostatic, Exchange Repulsion, Charge Transfer, Diffusion

FMO Fragment Series



$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$
$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$
$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

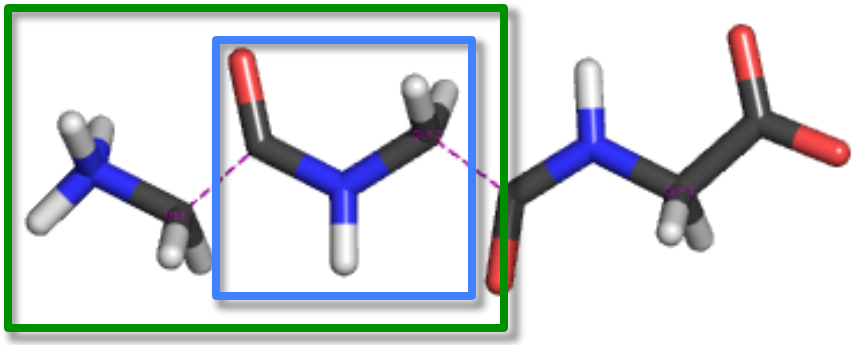
FMO Fragment Series



FMO1

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$
$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$
$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

FMO Fragment Series



FMO1

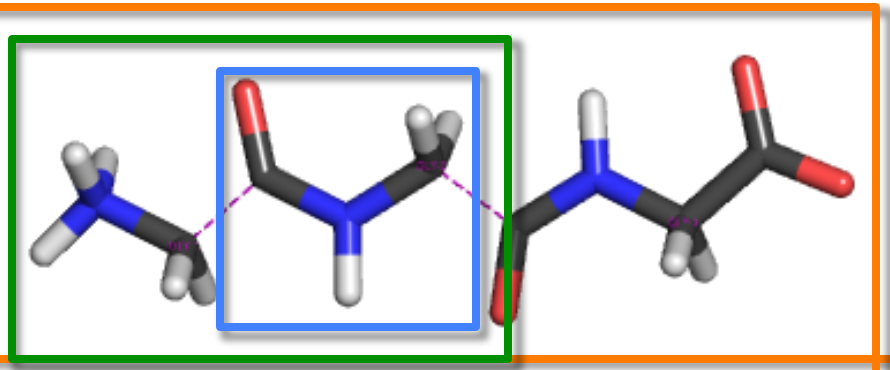
FMO2

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$

$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$

$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

FMO Fragment Series



FMO1

FMO2

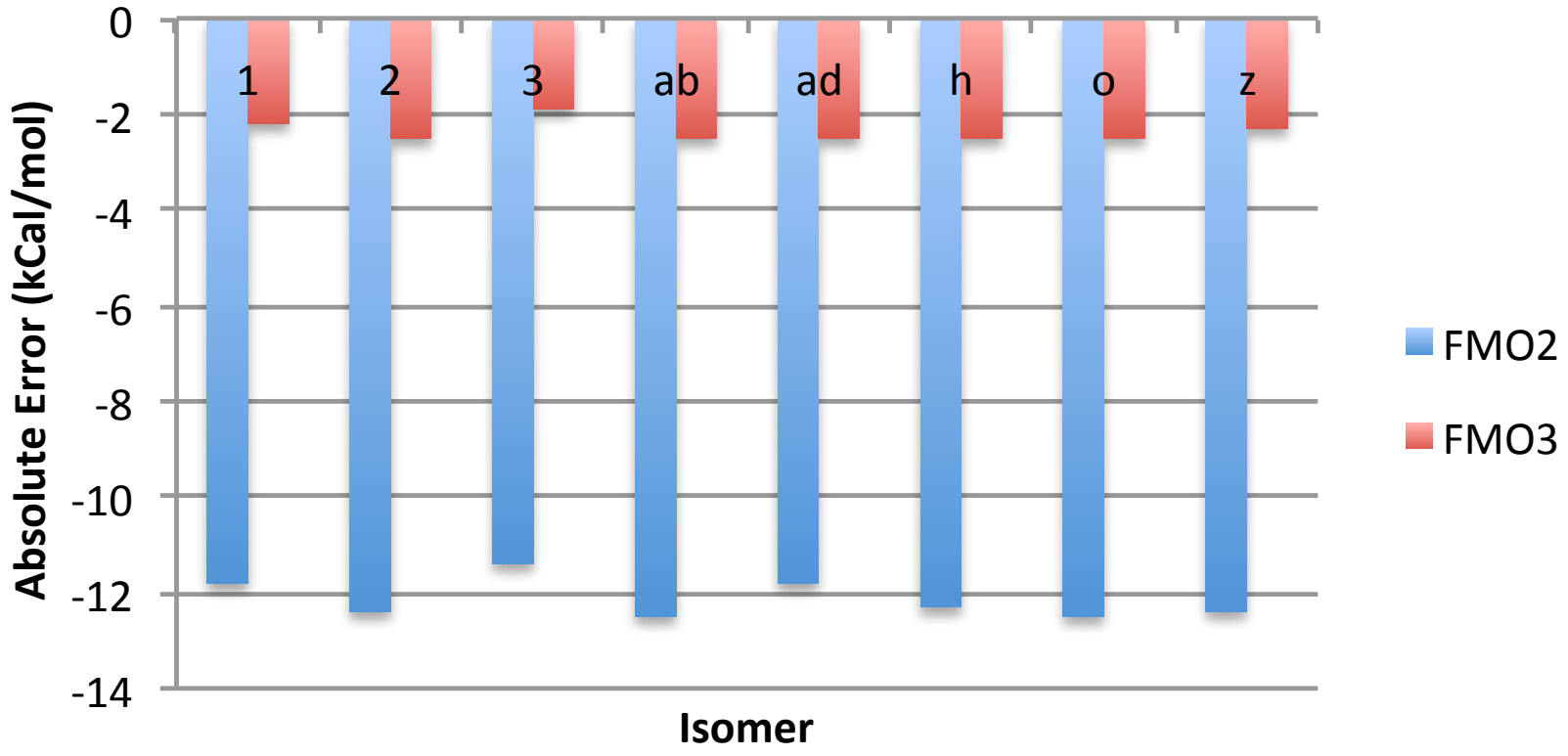
FMO3

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} (E_{IJ} - E_I - E_J)$$

$$+ \sum_{I>J>K}^{N} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J)$$

$$- (E_{JK} - E_J - E_K) - (E_{KI} - E_K - E_I)\} + \cdots,$$

FMO ‘Level’ Accuracy

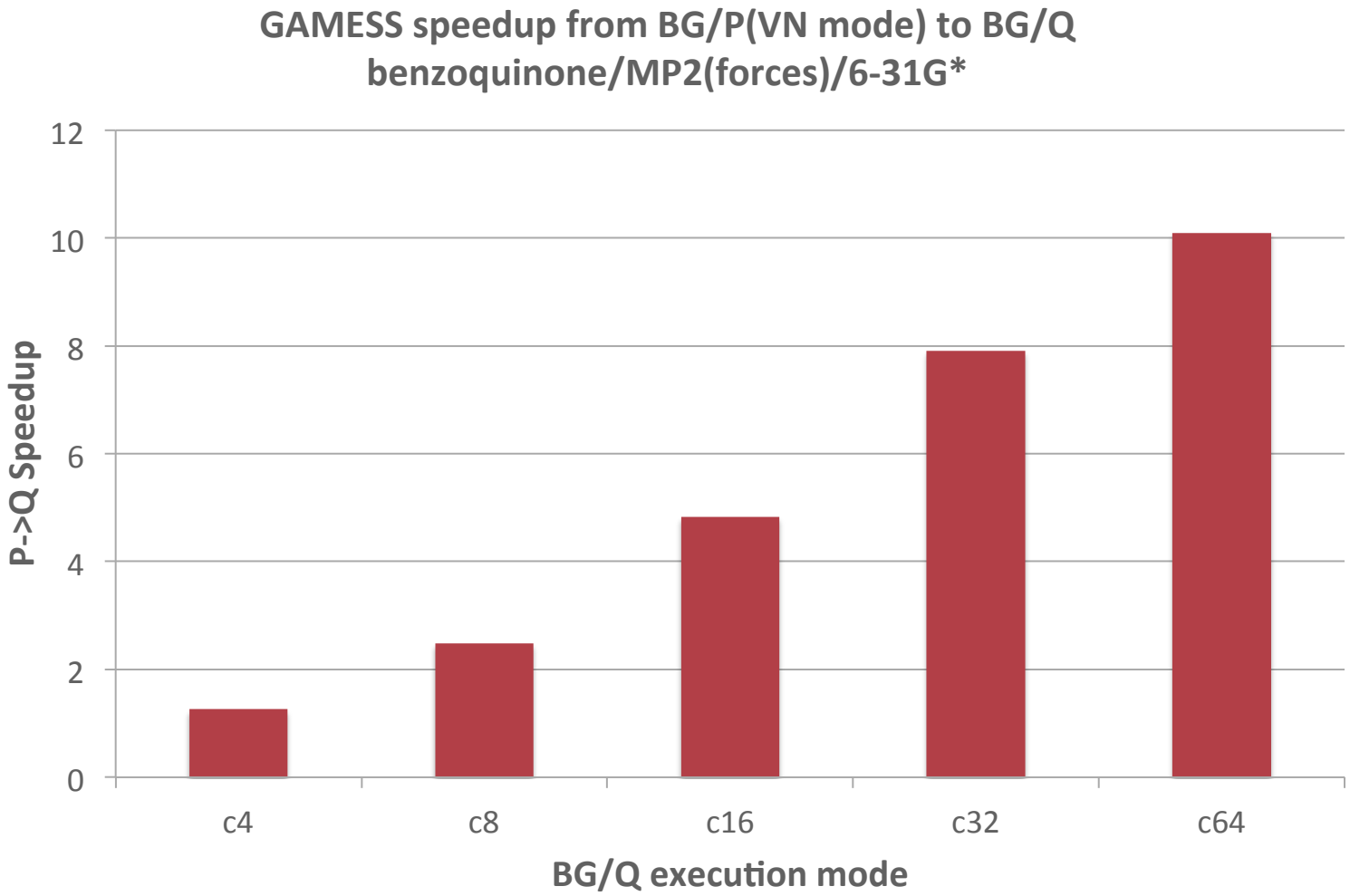


32 waters/FMO(2,3)/-MP2/6-31G*

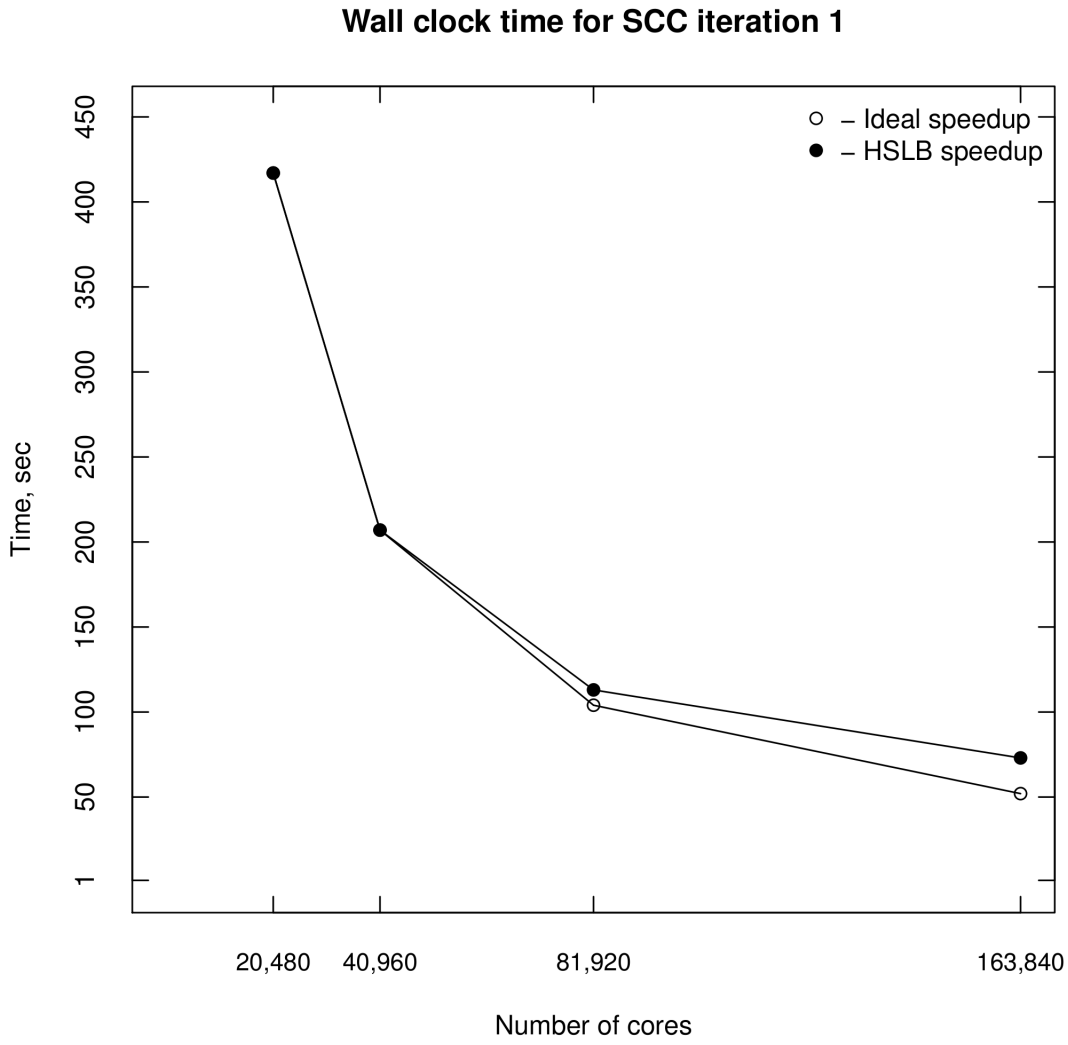
Porting GAMESS to Blue Gene/Q

- IBM XL compiler
- ESSL
- MPI-1

Performance on Mira

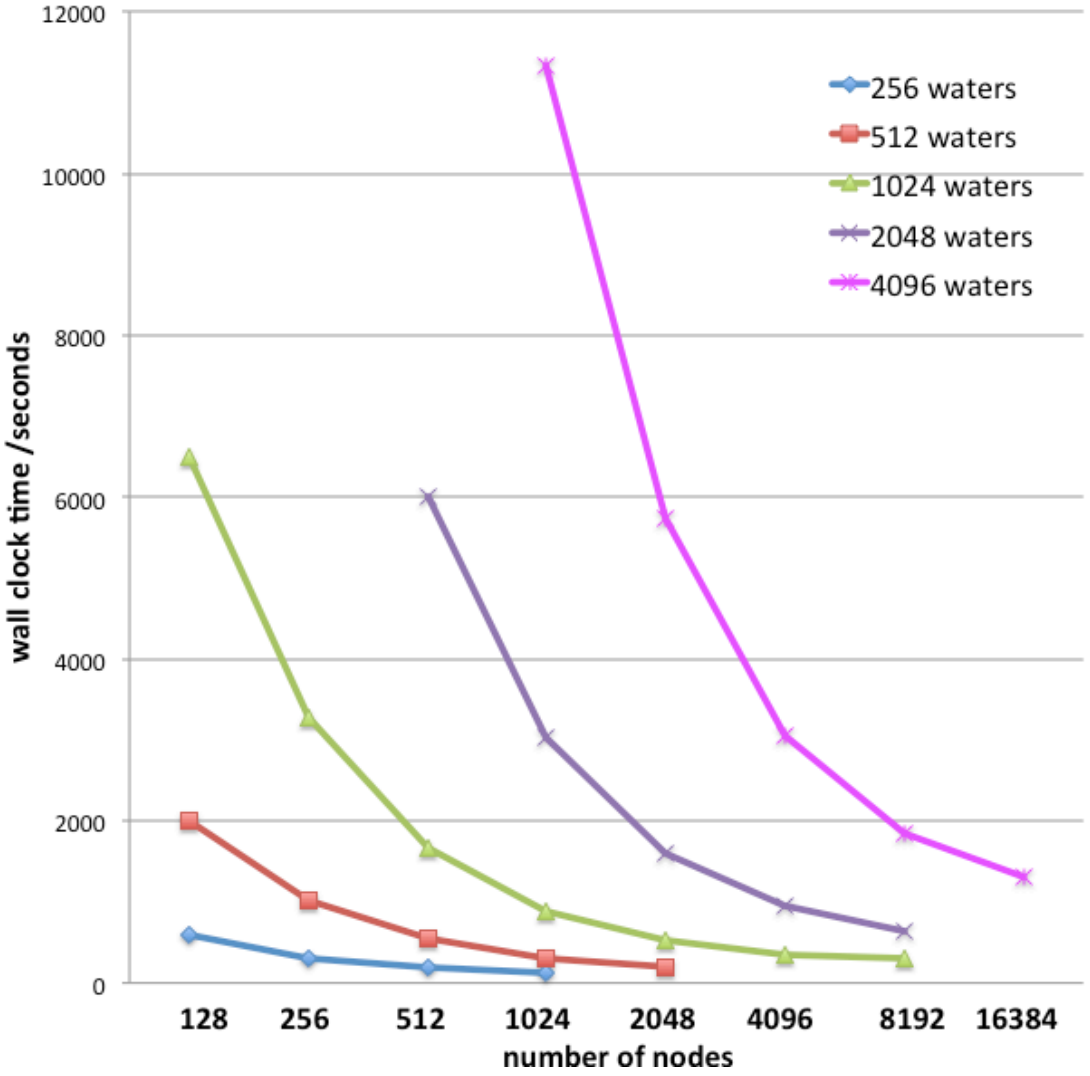


Scalability of FMO on Intrepid



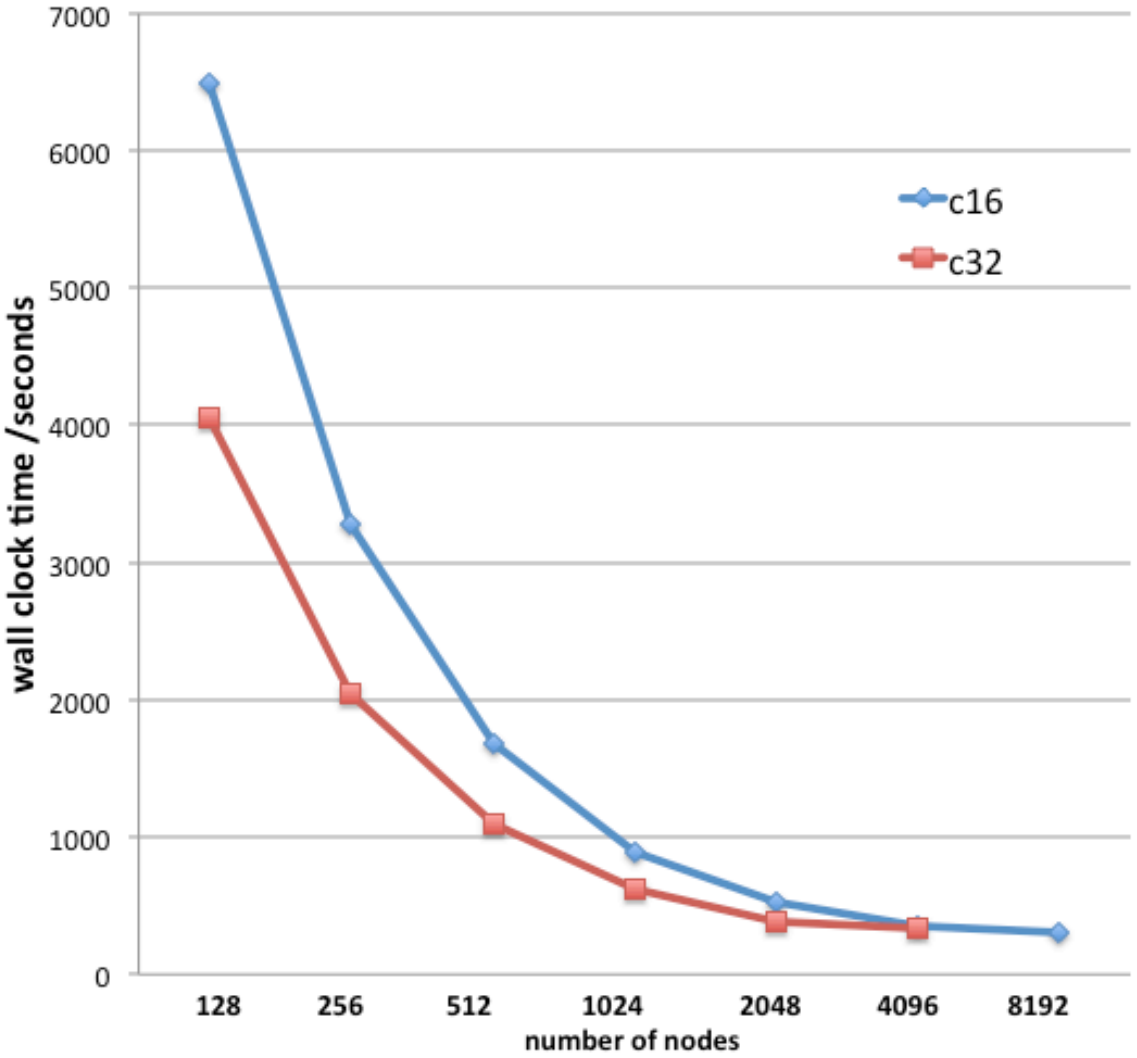
COX-1 complexed with ibuprofen (17,767 atoms) using FMO2-RHF/6-31G*

Scalability of FMO on Mira



FMO2-MP2(forces)/aug-cc-pvdz, c16 mode

Scalability of FMO on Mira



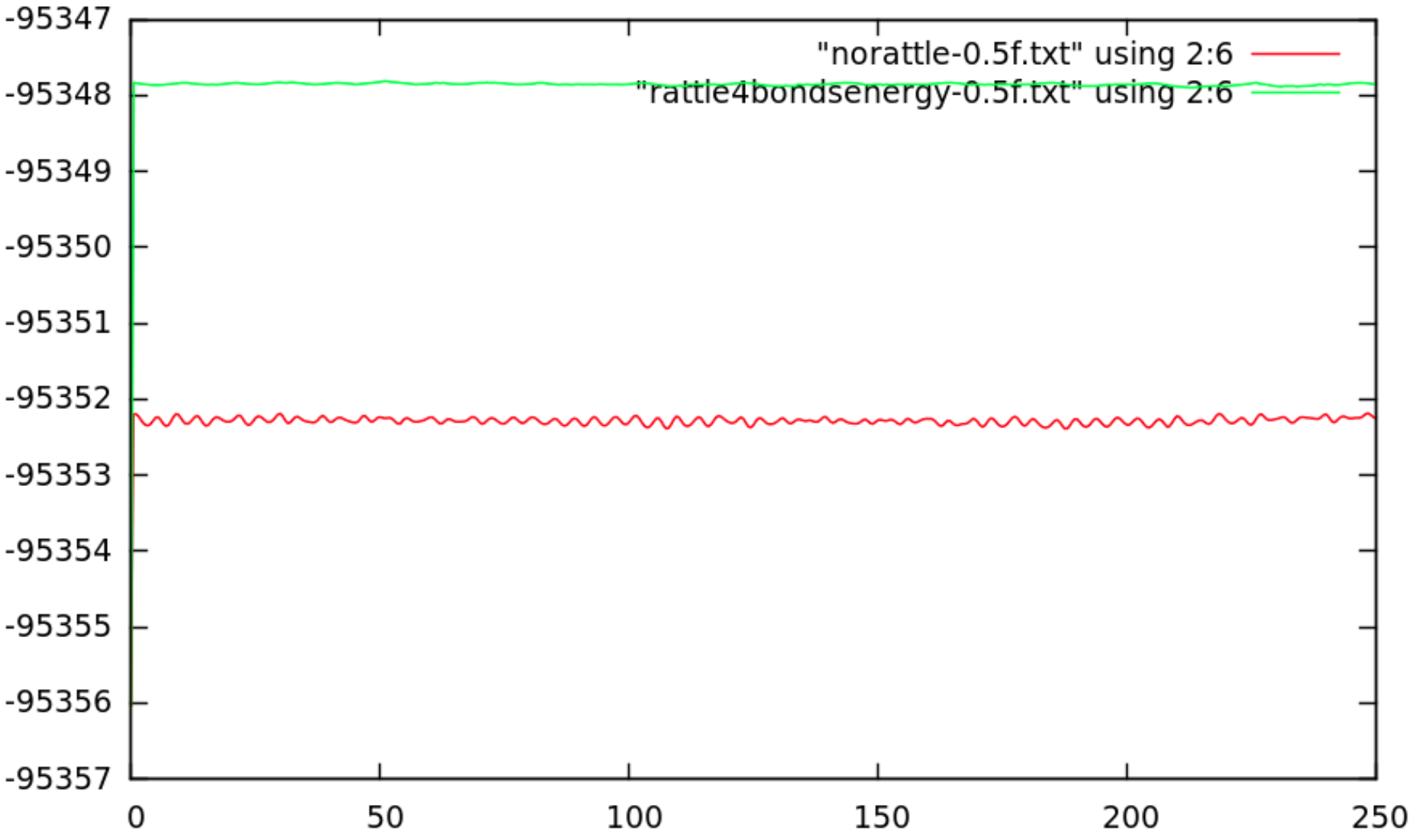
1024 waters/FMO2-MP2(forces)/aug-cc-pvdz

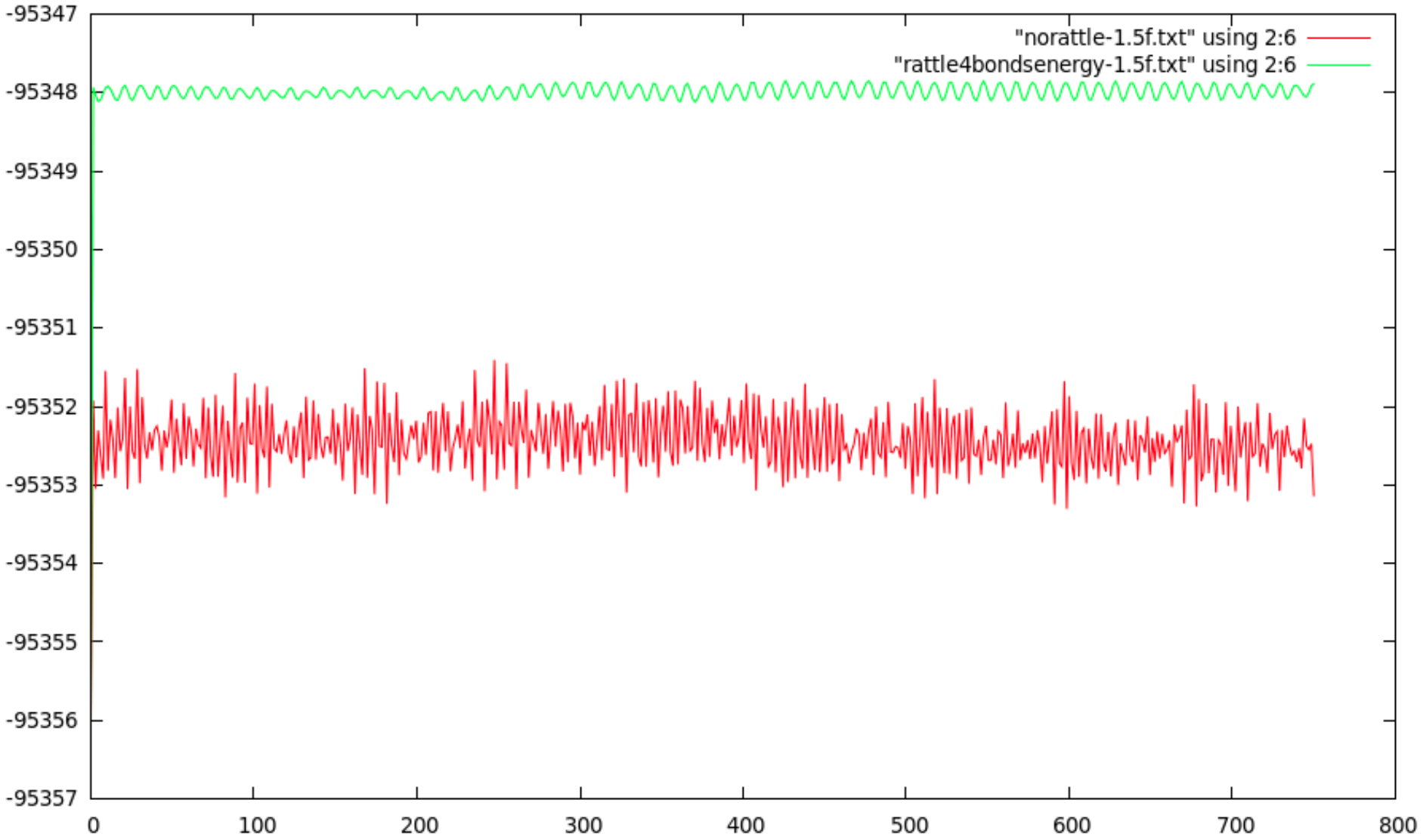
What is enabled by Mira over Intrepid

- This Early Science project focuses on bulk properties of water:
  - larger embedded clusters
  - longer dynamics simulations
  - higher levels of theory (accuracy)
  - larger basis sets

Code Modifications, New Algorithms: RATTLE

- A form of Constrained Dynamics to allow larger time steps and avoid local minima
- Constraints are applied to both distances and velocities simultaneously
- Lagrange-multiplier based method developed especially for velocity version of Verlet integrators
- Helps the trajectory to sample the most important regions of phase space
- Improves the stability of the trajectory by removing abrupt shifts in molecule positions


-95347
-95348
-95349
-95350
-95351
-95352
-95353
-95354
-95355
-95356
-95357
0
50
100
150
200
250
"norattle-0.5f.txt" using 2:6
"rattle4bondsenergy-0.5f.txt" using 2:6


"norattle-1.5f.txt" using 2:6
"rattle4bondsenergy-1.5f.txt" using 2:6
-95347
-95348
-95349
-95350
-95351
-95352
-95353
-95354
-95355
-95356
-95357
0
100
200
300
400
500
600
700
800

Summary

- Successfully ported GAMESS on Blue Gene/Q systems
- Successfully ran test and benchmark calculation on Intrepid and Blue Gene/Q systems
  - Relative node-to-node speedup of up to 10x
- Demonstrated scalability and performance of FMO
  - Scales up to 16 racks of Mira
- Implemented a new functionality on GAMESS (RATTLE algorithm) which would significantly reduce the computational cost of the project

Acknowledgment

- This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-ACO2-06CH11357.
- ALCF's Early Science Program